

# 프로젝트 보고서

## Project Report

항목	내용
개발 프로젝트명	AI 회귀 분석을 이용한 대한민국 인구 예측
프로젝트팀	4조 (송봉환, 김윤준, 채수병)
프로젝트 기획의도	<p>본 프로젝트는 대한민국의 고령화와 저출산 문제를 분석하고 예측하기 위한 빅데이터 분석 서비스입니다.</p> <p>현재 진행되고 있는 고령화 및 저출산으로 인한 심각한 문제를 인지하고, 미래 세대를 위하여 문제의 심각성을 인지하여야 합니다.</p> <p>미래 인구 예측을 바탕으로 인구 피라미드를 생성하고, 지역 소멸 위험도를 예측하여 적절한 정책과 대응 방안을 마련하는 것을 목표로 합니다.</p>
개발 목표 및 계획 수립	<ul style="list-style-type: none"><li>프로젝트 기간 : 2024년 12월 30일 ~ 2025년 1월 24일</li><li>예측 목표 : 2025년부터 2075년까지를 목표로 인구 예측을 시행합니다.</li><li>출생인구, 사망인구, 이주인구 데이터 등 여러 데이터를 활용.</li><li>데이터는 연도별로 제공되어야 하며, 과거 10년 이상의 데이터를 기본으로 합니다.</li><li>과거 데이터 분석을 통한 미래 인구를 예측합니다.</li></ul>
구현 가능 서비스	<ol style="list-style-type: none"><li>지역별 미래 인구 예측 서비스<ul style="list-style-type: none"><li>여러 데이터를 활용하여 지역별 출생아 및 사망 인구를 예측하여 지역별로 5년 단위의 미래 인구를 예측하는 서비스입니다.</li></ul></li><li>인구 피라미드 생성 서비스<ul style="list-style-type: none"><li>예측된 인구 데이터를 기반으로 인구 피라미드를 생성합니다.</li><li>사용자는 특정 연도를 선택하여 피라미드를 볼 수 있습니다.</li><li>피라미드는 시각적으로 보기 쉽고 이해하기 쉬운 형태로 제공됩니다.</li></ul></li><li>지역 소멸 위험도 예측 서비스<ul style="list-style-type: none"><li>지역별 인구 예측 데이터를 기반으로 지역별 인구 감소에 대한 지역 소멸 위험도를 예측합니다.</li><li>각 지역별로 위험도 등급을 수치 및 게이지 형식으로 구분하여 시각적으로 제공.</li><li>사용자는 5년 단위의 특정 연도를 선택하여 인구 소멸 위험도를 볼 수 있습니다.</li></ul></li></ol>
시스템 아키텍처	<ol style="list-style-type: none"><li>데이터 처리 및 예측 모델:<ul style="list-style-type: none"><li>Python을 사용하여 데이터 처리, 예측 모델을 구현.</li><li>Pandas, NumPy를 활용한 데이터 처리.</li><li>Script를 이용한 데이터 시각화.</li></ul></li><li>프론트엔드 / 백엔드를 이용한 웹 제공 및 서비스 시각화</li></ol>

서비스 제공 방식	웹 기반 서비스로 제공되며, 사용자는 원하는 서비스 선택, 제공을 기반으로 미래 인구 예측을 피라미드로 제공하며, 지도에 지역별 표시를 구현하여 원하는 연도 선택 후 해당 지역에 대한 미래 인구 및 지역 소멸 위험도 시각화 제공합니다.
-----------	---

데이터 분석	데이터 수집 목록		
	데이터 이름	데이터 목적	데이터 전처리
	가임 여성	출생 인구 input data	07~23년도 지역별 데이터
	사교육비	출생 인구 input data	07~23년도 지역별 데이터
	임금	출생 인구 input data	07~23년도 지역별 데이터
	집 값	출생 인구 input data	07~23년도 지역별 데이터
	혼인 건수	출생 인구 input data	07~23년도 지역별 데이터
	출생아	출생 인구 타겟 값	07~23년도 지역별 데이터
	노인 인구	사망 인구 input data	07~23년도 지역별 데이터
	암 환자	사망 인구 input data	07~23년도 지역별 데이터
	정신질환 환자	사망 인구 input data	07~23년도 지역별 데이터
	순환계질환 환자	사망 인구 input data	07~23년도 지역별 데이터
	사망 인구	사망 인구 타겟 값	07~23년도 지역별 데이터
	5세별 인구	인구 피라미드 구현을 초기 데이터	23년도 5세별 인구 데이터
	지역 인구	지역 소멸도 표현을 위한 데이터	23년도 지역별 인구 데이터

각 데이터를 도시별 년도별로 전처리하여 데이터 사용.



GridSearchCV 및 PolynomialFeatures 사용 또는 미사용하여 Ridge와 Lasso를 사용하여 트레이닝 값과 테스트 값 비교하였더니 Ridge에서 GridSearchCV 및 PolynomialFeatures를 사용했을 때의 값이 가장 이상적인 값이 확인됨.

구현한 머신 러닝으로 미래 지역별 인구수 예측 완료

시스템 구현	<ol style="list-style-type: none"> <li>인구 피라미드 생성 서비스 <ul style="list-style-type: none"> <li>5세별 인구를 10세별로 묶고, 예측된 출생인구를 피라미드 제일 밑부분에 입력, 예측된 사망 인구를 기존 데이터에 빼는 형식으로 10년 단위의 인구 피라미드 생성 완료</li> </ul> </li> <li>지역 소멸 위험도 예측 서비스 <ul style="list-style-type: none"> <li>년도마다 예측된 지역별 인구수를 5년 단위로 인구수 시각화 완료</li> <li>소멸도는 23년도 지역별 인구수를 기반으로  지역 소멸 위험도 = <math>100 - ((\text{예측된 인구수} / 23\text{년도 인구수}) \times 100)</math>으로  계산하여 시각화 함.  지역 소멸 위험도 표현하기 위해 게이지로 표현하여 제공함.</li> </ul> </li> </ol>
평가	<p>목표한 기능을 전부 수행하였지만, 그 과정에서 많은 어려움과 시행착오를 겪었습니다.</p> <p>GridSearchCV 및 PolynomialFeatures 사용 또는 미 사용하여 Ridge와 Lasso를 사용하여 트레이닝 값과 테스트 값 비교시에도 원하는 값이 정확하게 나오지 않아 어려움을 겪었습니다.</p> <p>이 부분에서 팀원과의 의사소통으로 문제점을 파악하고자 노력했으며, 찾아낸 결과는 데이터의 부족으로 판단하여 데이터를 더 수집하여 활용하고, 테스트 값을 높여서 어느 정도는 원하는 값을 얻을 수 있었습니다.</p> <p>계획을 세밀하게 짜고 시작하였지만, 프로젝트를 진행하다 보니 중간중간 생각대로 진행되지 않아 딜레이가 되기도 했지만 기한내에 마무리가 되었으며, 결과물도 만족스럽습니다.</p> <p>첫 머신 러닝으로 만든 결과물이다 보니 보았을 때 부실한 면이 보입니다. 더 많은 데이터의 활용으로 보다 정확한 기능을 만들어보고, 새로운 기능도 생각해 봐야 합니다.</p>